

Identification of Mango Leaf Diseases Using a Convolutional Neural Network

¹Kamal Gulati* and ²Jong Hyun Yang

Author's Affiliation

¹Amity University, Noida, Uttar Pradesh, India

²Chungwoon University, Incheon, South Korea

*Corresponding Author:

Kamal Gulati,

Amity University, Noida, Uttar Pradesh, India

E-mail: drkamalgulati@gmail.com,

<https://orcid.org/0000-0002-1186-1426>.

Keywords:

Mango,
Convolutional Neural
Network,
Diseases, Accuracy.

Received on 12.03.2026

Received on 22.04.2026

Accepted on 03.06.2026

Abstract

Mango cultivation contributes substantially to agricultural production in many regions, yet mango trees are affected by several diseases that can reduce crop yield and quality. Timely identification of these diseases is therefore important for effective disease management and sustainable production. In this study, a Convolutional Neural Network (CNN) is applied to identify diseases from mango leaf images. The developed CNN achieved an accuracy of 96.88% when evaluated using 2000 high-quality images obtained from the Mendeley dataset. The image collection represents Gall Midge, Powdery Mildew, Sooty Mould, and healthy leaves. The obtained accuracy, recall, and F1-score indicate that the model can distinguish diseased leaves with relatively few false positive and false negative classifications. The findings demonstrate the potential of deep-learning-based image analysis as a tool for supporting plant disease management and, consequently, agricultural productivity and food security.

How to cite this article: Gulati K. and Yang J.H. (2026). Identification of Mango Leaf Diseases Using a Convolutional Neural Network. *Bulletin of Pure and Applied Sciences-Botany*, 45B (1), 53-62.

1. INTRODUCTION

Mango (*Mangifera indica*) is an important fruit crop cultivated extensively for its desirable taste, nutritional value, and economic contribution. Mango fruit contains important nutrients, including vitamins A and C. Despite its importance, mango production is affected by several diseases that can reduce yield and compromise fruit quality. Such problems may arise from fungal, viral, and bacterial pathogens as well as unfavorable environmental conditions. India is the leading mango-producing country and accounts for more than 40% of global mango production. Mango is therefore an important crop within the Indian agricultural economy. Different diseases affected about 30% to 40% of agricultural production, and mango leaf diseases may

remain unnoticed because visual inspection is limited.

Limited awareness of mango leaf diseases among farmers can contribute to reduced fruit production. The effects of individual diseases on mango crops vary (Gulavnai & Patil, 2019). Common diseases and disorders of mango include gall infestation, Webber attack, mango deformation, stem miner, anthracnose, and *Alternaria* leaf spots. These problems may be associated with bacteria, viruses, fungi, parasites, and unfavorable environmental conditions. Leaf infection can interfere with photosynthesis and, in severe cases, contribute to plant decline. Disease symptoms and their distribution on the leaf can assist in identifying the affected condition. Gall Midge, Powdery Mildew, and Sooty Mould are particularly

important because they can reduce yield, lower product quality, and cause economic losses.

Mango gall midge: Gall midges are small insects occurring in many mango-growing regions. Their larvae develop within plant tissues and form abnormal structures known as galls. Infestation can affect leaves, flowers, fruit, and stems. Leaf gall midge is an important mango pest, and severe infestation can reduce inflorescence formation and subsequently lower fruit production. Gall leaves on trees have also been reported to carry anthracnose inoculum. Young larvae feed on newly emerging leaves and produce small, raised, wart-like galls, particularly on the underside of the leaf. Adults emerge from affected leaves after approximately two weeks during the summer season, leaving holes in the leaf tissue. Such damage can reduce photosynthetic activity, deform leaves, promote premature leaf drop, and ultimately affect fruit production.

Powdery Mildew: Powdery mildew, caused by the fungus *Oidium mangiferae*, is a common and potentially destructive disease of mango. Under severe infection, crop losses may exceed 50%. The disease can affect inflorescences, stems, young leaves, and immature fruits. A characteristic symptom is the formation of a white, powder-like fungal layer on infected plant parts. Conidia are produced in chains on conidiophores. Infected flowers may fail to open and can fall prematurely, while young fruits covered by fungal growth may also drop. Poor fruit set and substantial flower and fruit shedding can result in losses of 70–90% on an individual plant. Powdery mildew is recognized among agriculturally important fungal diseases, and fungal infection can reduce photosynthetic activity (Özçelik et al., 2026).

Sooty Mould: Sooty mould is a secondary fungal problem frequently associated with sap-feeding insects such as aphids, scales, and whiteflies. The fungi develop on the honeydew deposited by these insects and form a dark, velvety coating over leaves, stems, and fruit. Although sooty mould does not directly infect mango tissues in the same way as primary pathogens, its surface growth can interfere with light penetration and gas exchange, thereby affecting photosynthesis. Heavy deposits can also reduce fruit visibility

and contribute to quality loss and reduced consumer appeal.

Because these diseases are difficult to identify consistently through manual observation, automated image-based methods can support disease detection and diagnosis. Techniques such as K-means clustering and Support Vector Machine (SVM) have been investigated for disease-region categorization and segmentation (Srunitha & Bharathi, 2018). Deep-learning methods have subsequently been applied to classify mango-related images (Mettleq et al., 2019), while CNN-based semantic segmentation has been used to estimate the number of mangoes present in orchard plants (Kestur et al., 2019). Deep learning has also been applied to mango leaf disease diagnosis (Arivazhagan & Ligi, 2018). In the present study, a CNN-based approach is used for early identification of three target mango leaf diseases: gall midge, powdery mildew, and sooty mould. Machine learning is also increasingly used in precision agriculture for pest and disease detection and for data-driven agricultural decision-making (Taqvi, 2025).

2. RELATED WORK

Recent studies have introduced a range of methods for automated plant-leaf disease detection and classification. Many of these approaches employ deep convolutional networks to extract image features and perform classification, with applications extending to plant, animal, and fish disease diagnosis (Cho et al., 2024; AlZubi, 2023; Wasik and Pattinson, 2024) as well as industrial applications (Porwal, 2024). Recent crop-leaf studies also support the use of deep-learning image classification for automated disease recognition. For example, Tholupuluri et al. (2024) applied an EfficientNetB0-based method together with data augmentation for rice leaf disease detection.

Safdar et al. (2019) used watershed-based segmentation to separate diseased regions from citrus leaf images and reported 95.5% segmentation accuracy across three citrus datasets. Febrinanto et al. (2019) investigated K-nearest-neighbour-based segmentation and reported a detection rate of 90.83%. Their two-cluster and nine-cluster approaches for identifying diseased regions in citrus leaves

achieved 99.17% accuracy under the reported minimum bond parameter of 3%. Adeel et al. (2019) reported 90% and 92% accuracy for segmentation and classification of grape images from the PlantVillage dataset. Their LCHR enhancement procedure incorporated LAB colour transformation to select an appropriate channel for thresholding. Bhargava et al. (2020) classified fruit images into green, red, and blue categories and applied a split-and-merge procedure to reduce background noise.

Ali et al. (2022) investigated mango leaf disease detection and provided standardized mango leaf images for machine-learning training and testing through the Mendeley Data repository (Ali et al., 2022; Ahmed et al., 2023). The dataset can support subsequent work on mango disease recognition and LeafNet development. Rizvee et al. (2023) developed LeafNet, a CNN designed to diagnose seven major mango leaf diseases. The reported results demonstrate the applicability of specialized CNN architectures to agricultural disease identification.

Li et al. (2020) used grid search to examine different combinations of model hyperparameters and identify a suitable configuration. Andrushia et al. (2019) proposed an automated diagnostic approach for mango skin disease in which texture, colour, and shape features were selected from images using artificial bee colony optimization and then classified with SVM. Raghavendra et al. (2021) reported an efficient wavelength-selection

procedure for mango fault detection with an accuracy of 84.5%. Kumari et al. (2021) developed an automated Mango Defect Identification System based on Enhanced Fuzzy K-means clustering, maximal correlation analysis of principal components, and a back-propagation-based discriminative classifier. Patel et al. (2021) reported a grayscale machine-vision system for mango fault detection with an accuracy of 97.88%.

3. MATERIAL AND METHODS

Deep learning uses multilayer neural networks to learn complex patterns from data. In image-processing applications, convolutional neural networks (CNNs) are commonly used because they can learn spatial features directly from images. In this study, images of healthy and diseased mango leaves are used to train a CNN for classification. The methodology comprises four main stages: data collection, model development, model training, and model testing.

Dataset

A suitable dataset is required throughout the recognition workflow, from model training through performance evaluation. The training data used in this work were obtained from the Mendeley database. The dataset contains 2000 mango leaf photographs distributed among four categories: gall midge, powdery mildew, sooty mould, and healthy (Figure 1).

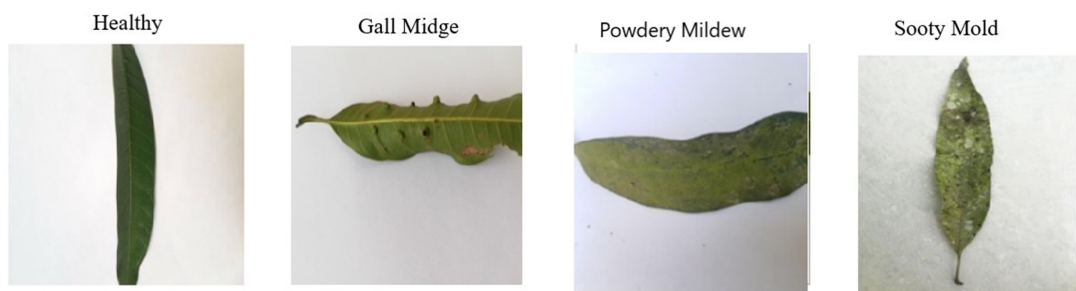


Figure 1: Images of mango plant leaves categorized as Healthy, Gall Midge, Powdery Mildew, and Sooty Mould.

3.2 Image pre-processing and labelling

Before CNN training, the original images were subjected to image pre-processing. Because images collected from different sources can have

different dimensions, they were resized and rescaled to provide a consistent input format. Standardizing the image dimensions also helps maintain computational consistency during

model training. The photographs were resized to $256 \times 256 \times 3$ and converted to grayscale. The processed mango leaf images were then reviewed and assigned the corresponding health or disease label. Four classes were used in the training and testing datasets: healthy, gall midge, powdery mildew, and sooty mould.

3.3. Model Architecture

The CNN model described in this study is intended for binary classification with two classes. Its architecture contains six convolutional layers (Conv2D), with increasing filter and kernel configurations based on 3×3 kernels, followed by six max-pooling

(MaxPooling2D) stages for spatial downsampling. After the feature maps are flattened, the network passes the information to a dense layer containing 64 units with ReLU activation. A final dense layer with softmax activation produces the class probabilities. This arrangement follows a conventional image-classification structure in which convolutional layers learn image features and fully connected layers model nonlinear relationships. Depending on the characteristics of the dataset, parameters such as the number of layers and filter sizes can be adjusted. The principal CNN architecture and training settings are listed in Table 1

Table 1: Convolutional Neural Network (CNN) Hyperparameters.

Parameter	Description
Image size	256×256 pixels
Conv (2D) Layer	6
Max pooling (2D) Layer	6
Dropout Rate	0.5
Function of Activation	"ReLU"
Epochs	25
Size of Batch	32

The sequential CNN contains six convolutional layers that extract increasingly complex features from the input using 32 or 64 filters. Six max-pooling layers reduce the spatial dimensions of the feature maps. Dropout is applied at a rate of

0.5 as a regularization mechanism to support model generalization. The network uses ReLU activation, while training is performed for 25 epochs with a batch size of 32. Figure 2 presents the overall CNN architecture.

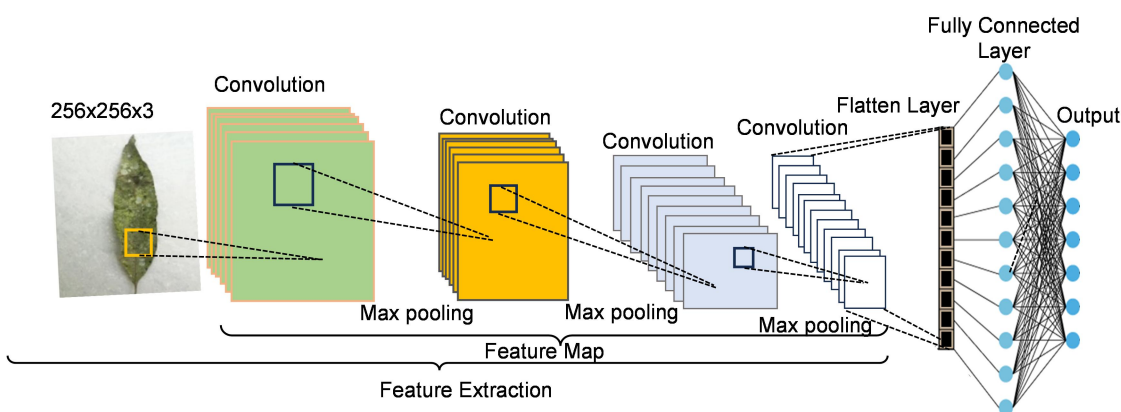


Figure 2: CNN model Architecture.

The CNN procedure is further represented using the mathematical expressions below. The input image is represented by coordinates (x, y) and has dimensions $N \times M$. The first equation is used to determine the image histogram.

$$h_i(k) = O_j \quad (1)$$

where $h_f(k)$ denotes the image histogram, f represents the frequency of occurrence, and O_j ($j = 1, 2, 3, \dots, k-1$) represents the grayscale events. Based on this expression, the range of infected pixels is represented by:

$$\sim h_f(k) = h_f(k)[I_{j1}]_{k_1, k_n}, \quad (2)$$

where J represents pixel values and I denotes the affected area. The infected region extends from k_1 to k_n , and the complete affected region is represented by $\sim h_f(k)$. The image-variable quantity in Equation 3 is calculated as:

$$\sigma = \frac{1}{MN} \sum_{i=0, j=0}^{N-1, M-1} (\xi_{ij})^2 - \mu^2 \quad (3)$$

The mean value is expressed as:

$$\mu = \sum_{i=0, j=0}^{N-1, M-1} (\xi_{ij}) \times \frac{1}{MN} \quad (4)$$

The weight and bias terms of the convolutional layer are expressed as:

$$S = \sum \xi xy + w + b \quad (5)$$

where W denotes the weight matrix of the i th layer, b denotes the bias matrix, and S represents the characteristics of the first convolutional layer. The variables x and y refer to the processed image. A ReLU activation layer is then applied. The subsequent convolutional layer uses a [3, 3] filter, a [1, 1] stride, and 64 channels/filters. ReLU activation is used for this layer. A max-pooling operation follows with a [2, 2] filter and [2, 2] stride. The complete model uses six convolutional and max-pooling stages together with dropout, flattening, and dense layers.

3.4. Evaluation Metrics

The CNN is evaluated using accuracy, precision, recall, and F1-score. These measures are calculated from the numbers of true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) classifications.

Accuracy

Accuracy represents the proportion of correctly classified samples, whether positive or negative, among all evaluated samples.

$$Accuracy\% = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

Precision

Precision measures the proportion of correct positive predictions among all samples predicted as positive.

$$Precision\% = \frac{TP}{TP + FP} \times 100$$

Recall

Recall represents the proportion of correctly identified positive samples among all actual positive samples.

$$Recall\% = \frac{TP}{TP + FN} \times 100$$

F1 Score

The F1-score combines precision and recall into a single measure and is useful when both types of classification performance need to be considered together.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Recall + Precision}$$

3. RESULTS AND DISCUSSION

The proposed CNN follows a conventional image-classification structure in which convolutional layers extract visual information and fully connected layers perform classification. Max-pooling operations progressively reduce the spatial dimensions of the feature maps while retaining important image characteristics. The final dense layer applies softmax activation to generate class probabilities. Model performance depends on the properties of the dataset, the selected hyperparameters, and the classification task. Different architectural configurations and parameter settings may therefore be examined during model development.

The network was trained for 25 epochs, corresponding to 25 complete passes through the training dataset. Figure 3 presents the

training and validation accuracy together with the associated loss values. The reported loss was 0.1056, representing the difference between predicted and actual values. Training accuracy reached 96.15%. For the CNN model, the validation loss was 0.1507 and validation accuracy was 94.27%. Evaluation using the

independent validation data indicates that the model was able to apply the learned image patterns to previously unseen samples. Although the validation accuracy was lower than the training accuracy, the results demonstrate the model's ability to identify mango leaf diseases from images.

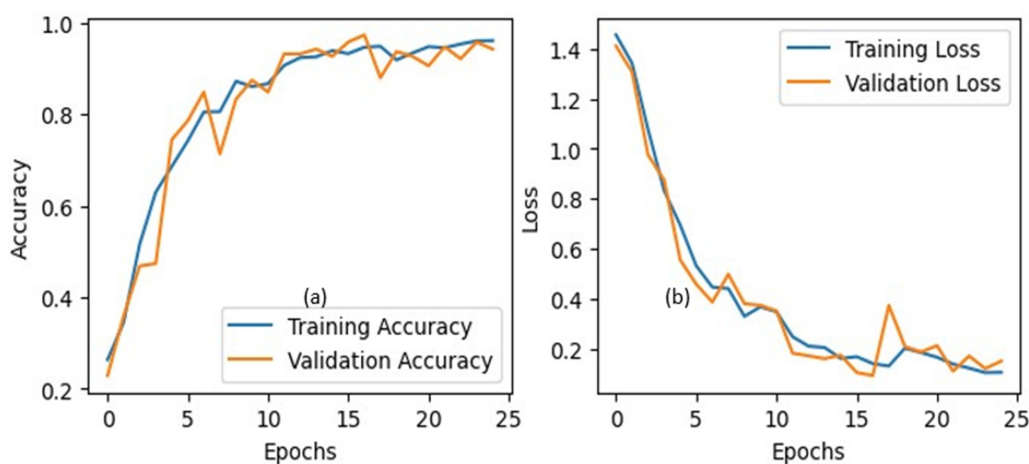


Figure 3: The accuracy and loss of the training and validation datasets.

Figure 4 presents the model predictions together with their confidence scores for multiple samples. The predictions show consistent classification across the disease categories.

Differences in confidence values provide additional information about how strongly the model separates the individual classes.

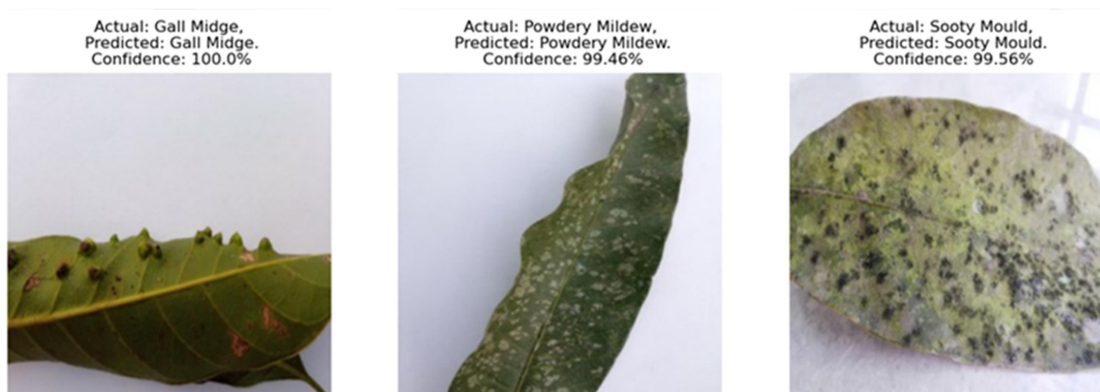


Figure 4: Actual and expected diseased images using the proposed CNN model.

The confusion matrix summarizes the classification performance for gall midge, powdery mildew, sooty mould, and healthy

leaves. The rows represent the actual class labels, whereas the columns represent the predicted labels (Figure 5).

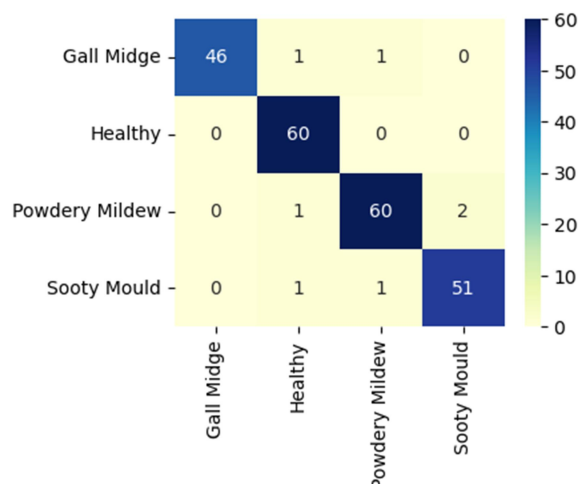


Figure 5: Confusion matrix.

The classification report summarizes the model performance for each of the four categories. Precision, recall, and F1-score are provided in Table 2. The reported precision values range from 1.0000 to 0.9623, recall ranges from 0.9583

to 0.9623, and F1-scores range from 0.9787 to 0.9623. These values indicate consistent classification performance across the four classes.

Table 2: Classification Accuracy, Precision, F1 Score, and Support

Class	Precision	Recall	F1 Score	Support
Gall Midge	1	0.9583	0.9787	48
Healthy	0.9524	1	0.9756	60
Powdery Mildew	0.9677	0.9524	0.96	63
Sooty Mould	0.9623	0.9623	0.9623	53
accuracy			0.9688	224
macro avg	0.9706	0.9682	0.9691	224
weighted avg	0.9692	0.9688	0.9687	224

The overall model accuracy was 0.9688, indicating that approximately 96.88% of the evaluated cases were assigned to the correct class. Recall, precision, accuracy, and F-score were used as the principal performance indicators. Macro-average and weighted-average F1-scores were also calculated to account for the distribution of samples among the classes.

For the ROC analysis, Sooty Mould produced an AUC of 0.65, while Powdery Mildew produced an AUC of 0.76 (Figure 6). The Healthy class had an AUC of 0.75. These AUC values describe the model's ability to distinguish between the relevant classes in the ROC analysis. The curves for gall midge, powdery mildew, sooty mould, and healthy samples provide a class-wise view of the model's discrimination performance.

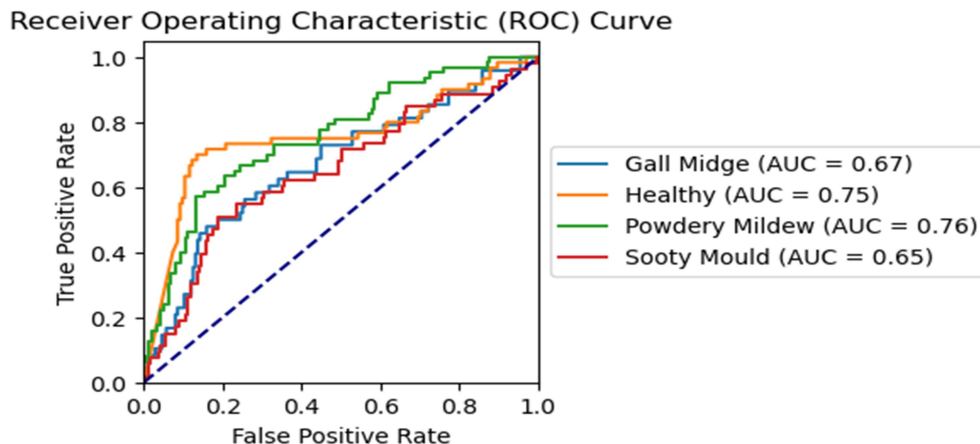


Figure 6: Receiver Operating Characteristic Curve for all four classes.

4. CONCLUSION

The study reports that the CNN approach achieved an overall accuracy of 96.88% for mango leaf disease identification. The use of deep-learning methods with a large image dataset provides an image-based approach for early recognition of the selected mango leaf diseases. Early identification of infected leaves may help farmers respond to disease problems before they spread further and may assist in reducing crop losses through appropriate management measures. The approach can also support more targeted disease-management practices and potentially reduce unnecessary use of broad-spectrum insecticides. The findings illustrate how machine-learning methods can be applied to recurring agricultural disease-management challenges. Since mango production is affected by multiple diseases, continued development of machine-learning-based diagnostic systems may contribute to improved disease monitoring and long-term crop management. Plant-derived antifungal compounds have also been investigated as environmentally friendly alternatives for managing phytopathogenic diseases.

5. LIMITATIONS AND FUTURE WORK

This study considers three principal diseases affecting mango leaves; however, mango plants can be affected by additional diseases that may involve leaves as well as stems and roots. Future investigations should therefore include a broader range of diseases and consider symptoms occurring in different parts of the

mango plant. The study also uses a fully connected layer classifier within the model. Future work may compare the performance of the current approach with hybrid machine-learning methods, including support vector machines, decision trees, and random forests. Such comparisons could provide further information about mango disease classification and may support improvements in diagnostic performance. The main objective of the present work is the recognition and classification of diseases affecting mango leaves. Diseases occurring in other plant species were outside the scope of the study. Future research should examine a broader set of plant diseases to determine how the developed approach can be adapted to other plant species.

REFERENCES

- Adeel, A., Khan, M. A., Sharif, M., Azam, F., Shah, J. H., Umer, T., & Wan, S. (2019). Diagnosis and recognition of grape leaf diseases: An automated system based on a novel saliency approach and canonical correlation analysis based multiple features fusion. *Sustainable Computing: Informatics and Systems*, 24, 100349. <https://doi.org/10.1016/j.suscom.2019.08.002>
- Ahmed, S. I., Ibrahim, M., Nadim, M., Rahman, M. M., Shejunti, M. M., Jabid, T., & Ali, M. S. (2023). MangoLeafBD: A comprehensive image dataset to classify diseased and healthy mango leaves.

- Data in Brief*, 47, 108941. <https://doi.org/10.1016/j.dib.2023.108941>
- Ali, S., Ibrahim, M., Ahmed, S. I., Nadim, M., Rahman, M., Shejunti, M. M., & Javid, T. (2022). MangoLeafBD Dataset [Data set]. *Mendeley Data*, Version 1. <https://doi.org/10.17632/hxsnvwt3r.1>
- AlZubi, A. A. (2023). Artificial Intelligence and its Application in the Prediction and Diagnosis of Animal Diseases: A Review. *Indian Journal of Animal Research*, 57(10), 1265-1271. <https://doi.org/10.18805/IJAR.BF-1684>
- Arivazhagan, S., & Ligi, S. V. (2018). Mango leaf diseases identification using convolutional neural network. *International Journal of Pure and Applied Mathematics*, 120(6), 11067-11079. Source: <http://www.acadpubl.eu/hub/>
- Bhargava, A., & Bansal, A. (2020). Automatic detection and grading of multiple fruits by machine learning. *Food Analytical Methods*, 13(3), 751-761. <https://doi.org/10.1007/s12161-019-01690-6>
- Cho, O. H., Na, I. S., & Koh, J. G. (2024). Exploring Advanced Machine Learning Techniques for Swift Legume Disease Detection. *Legume Research*, 47(7), 1221-1227. <https://doi.org/10.18805/LRF-789>
- Diana Andrushia, A., & Trephena Patricia, A. (2019). Artificial bee colony based feature selection for automatic skin disease identification of mango fruit. In *Nature Inspired Optimization Techniques for Image Processing Applications* (pp. 215-233). Springer. https://doi.org/10.1007/978-3-319-96002-9_9
- Febrinanto, F. G., Dewi, C., & Triwiratno, A. (2019). The implementation of k-means algorithm as image segmenting method in identifying the citrus leaves disease. *IOP Conference Series: Earth and Environmental Science*, 243(1), 012024. <https://doi.org/10.1088/1755-1315/243/1/012024>
- Gulavnai, S., & Patil, R. (2019). Deep learning for image-based mango leaf disease detection. *International Journal of Recent Technology and Engineering*, 8(3S3), 54-56. <https://doi.org/10.35940/ijrte.C1030.1183S319>
- Kestur, R., Meduri, A., & Narasipura, O. (2019). MangoNet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard. *Engineering Applications of Artificial Intelligence*, 77, 59-69. <https://doi.org/10.1016/j.engappai.2018.09.011>
- Kumari, N., Bhatt, A. K., Dwivedi, R. K., & Belwal, R. (2021). Hybridized approach of image segmentation in classification of fruit mango using BPNN and discriminant analyzer. *Multimedia Tools and Applications*, 80(4), 4943-4973. <https://doi.org/10.1007/s11042-020-09747-z>
- Li, H., Chaudhari, P., Yang, H., Lam, M., Ravichandran, A., Bhotika, R., & Soatto, S. (2020). Rethinking the hyperparameters for fine-tuning. *arXiv:2002.11770*. <https://arxiv.org/abs/2002.11770>
- Mettlegg, A. S. A., Dheir, I. M., Elsharif, A. A., & Abu-Naser, S. S. (2020). Mango classification using deep learning. *International Journal of Academic Engineering Research*, 3(12), 22-29. <https://ijeais.org/wp-content/uploads/2019/12/IJAER191205.pdf>
- Özçelik, H., Sağlam, B., & Öcal, A. S. (2026). New natural methods for combating mold fungi in agricultural products—The Isparta example. *Bulletin of Pure and Applied Sciences-Botany*, 45B(2), 1-17. <https://bpasjournals.com/botany/index.php/journal/article/view/174>
- Patel, K. K., Kar, A., & Khan, M. A. (2021). Monochrome computer vision for detecting common external defects of

- mango. *Journal of Food Science and Technology*, 58(12), 4550-4557. <https://doi.org/10.1007/s13197-020-04939-9>
- Porwal, S., Majid, M., Desai, S. C., Vaishnav, J., & Alam, S. (2024). Recent advances, challenges in applying artificial intelligence and deep learning in the manufacturing industry. *Pacific Business Review (International)*, 16(7), 143-152. Source: https://www.pbr.co.in/2024/2024_month/January/13.pdf
- Raghavendra, A., Guru, D. S., & Rao, M. K. (2021). Mango internal defect detection based on optimal wavelength selection method using NIR spectroscopy. *Artificial Intelligence in Agriculture*, 5, 43-51. <https://doi.org/10.1016/j.aiia.2021.01.005>
- Rizvee, R. A., Orpa, T. H., Ahnaf, A., Kabir, M. A., Rashid, M. R. A., Islam, M. M., ... & Ali, M. S. (2023). LeafNet: A proficient convolutional neural network for detecting seven prominent mango leaf diseases. *Journal of Agriculture and Food Research*, 14, 100787. <https://doi.org/10.1016/j.jafr.2023.100787>
- Safdar, A., Khan, M. A., Shah, J. H., Sharif, M., Saba, T., Rehman, A., ... & Khan, J. A. (2019). Intelligent microscopic approach for identification and recognition of citrus deformities. *Microscopy Research and Technique*, 82(9), 1542-1556. <https://doi.org/10.1002/jemt.23320>
- Srunitha, K., & Bharathi, D. (2018). Mango leaf unhealthy region detection and classification. In *Computational Vision and Bio Inspired Computing* (pp. 422-436). Springer International Publishing. https://doi.org/10.1007/978-3-319-71767-8_35
- Taqvi, S. (2025). Data-driven agricultural development in Assam through machine learning. *Journal of Artificial Intelligence in Agriculture and Applied Research*, 1(1), 34-48. <https://agrpjournals.com/journals/journal/index.php/jaiaar/article/view/7>
- Tholupuluri, G. B., Charith, B., & Gopal Krishna, U. M. (2024). Deep learning algorithms for rice leaf disease detection. *Bulletin of Pure and Applied Sciences-Zoology*, 43A(2S), 331-343. <https://bpasjournals.com/zoology/index.php/journal/article/view/243>
- Wasik, S., & Pattinson, R. (2024). Artificial Intelligence Applications in Fish Classification and Taxonomy: Advancing Our Understanding of Aquatic Biodiversity. *FishTaxa*, 31, 11-21. <https://doi.org/10.5281/zenodo.10851134>